

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-171396
 (43)Date of publication of application : 30.06.1997

(51)Int.Cl. G10L 5/04
 G06F 3/16
 G06F 17/21
 G10L 3/00

(21)Application number : 08-213255 (71)Applicant : BAISERA:KK
 (22)Date of filing : 25.07.1996 (72)Inventor : KATO KEIICHI
 KURITA HIROSHI

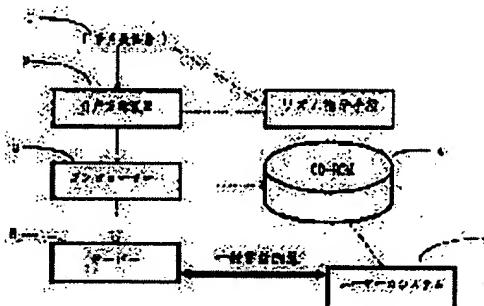
(30)Priority
 Priority number : 07293822 Priority date : 18.10.1995 Priority country : JP

(54) VOICE GENERATING SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To enable a user to select arbitrary and various synthetic voice generation by generating voice dictionary based on human voice, storing the dictionary in a hard recording medium and supplying it to the user.

SOLUTION: The result of the reading of a voice provider 1 is recorded in a voice recording device 2. Based on the result, a voice dictionay is generated. Then, the result is stored in a computer 3 and the dictionary is stored in a CD-ROM 4. The CD-ROM 4 is supplied to the user, and the dictionary, which is stored in the CD-ROM 4, is installed in a user's personal computer 5. Moreover, the dictionary and an information table are stored in a server 6 and the user connects the computer 5 to the server 6 through a public telephone line and may install an arbitrary dictionary in the user's personal computer 5. Based on the text analysis obtained from a manuscript, the user refers to a voice element dictionary, connects various voice elements complements the connecting sections and generates phoneme sequence. The rhythm generated by referring to a phoneme sequence and a rhythm dictionary, are synthesized to produce synthesis data.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-171396

(43)公開日 平成9年(1997)6月30日

(51)Int.Cl. ⁶	識別記号	序内整理番号	F I	技術表示箇所
G 10 L 5/04			G 10 L 5/04	F
G 06 F 3/16	330		G 06 F 3/16	330 K
	17/21		G 10 L 3/00	H
G 10 L 3/00			G 06 F 15/20	568 Z

審査請求 未請求 請求項の数41 FD (全 18 頁)

(21)出願番号	特願平8-213255	(71)出願人	394019820 有限会社バイセラ 神奈川県川崎市宮前区馬綱1634-1 コスモ宮崎代アバンシード305号
(22)出願日	平成8年(1996)7月25日	(72)発明者	加藤 圭一 神奈川県川崎市宮前区馬綱1634-1 コスモ宮崎台アバンシード305号
(31)優先権主張番号	特願平7-293822	(72)発明者	栗田 洋 神奈川県鎌倉市西鎌倉4-16-15
(32)優先日	平7(1995)10月18日		
(33)優先権主張国	日本 (J P)		

(54)【発明の名称】 音声発生システム

(57)【要約】

【目的】 音声合成においてユーザーが任意の多様な種類の音声の中から選択するためにユーザーに多様な種類の音声を辞書の形で供給しこの辞書を利用してユーザーの任意の種類の音声合成をできることを可能にするシステム。

【構成】 パソコンシステム、NCにおいて人の声を基にした音声辞書をユーザーに供給し、ユーザーのシステムにおいては、自由にこの辞書の各部分を組み合わせられるようし、ユーザーの任意の音声の音声合成を可能とする。ユーザーへの音声辞書の供給はハード媒体あるいは通信回線を通じて行う。

間の声を基にした音声辞書をユーザーの任意の組み合わせで選べること、前記人間の声を基にした音声辞書から選択した音声辞書の各要素の中から必用に応じて前記音声の可変可能部分の少なくとも1つをユーザーの任意の選択で可変させることを可能とすることを特徴とする特許請求範囲第1項記載および第2項記載の音声発生システム。

【請求項4】前記システムは入力テキストが予めテキストである情報、デジタル入力装置によって入力されテキストに変換された情報、OCRによって活字からテキストに変換された情報の少なくとも1つの情報から音声を発生させることを特徴とする特許請求範囲第1項記載および第2項記載の音声発生システム。

【請求項5】前記システムは入力がキーボード入力、活字文字の認識、手書き文字の認識入力から選ばれる少なくとも1つの方法の入力法であり、1文字入力の後の音声出力、単語確定後の音声出力、文節確定後の音声出力、文確定後の音声出力、1文字認識の後音声出力、単語認識後の音声出力、文節認識後の音声出力、文認識後の音声出力の少なくとも1つをユーザーの必用に応じて選択でき、前記1文字入力の後音声出力の場合にはアルファベット、ひらがな、及びカタカナから選ばれる少なくとも1つの方法で入力された1つの日本語文字を認識した直後に前記認識した文字に対応した音声を合成すること、単語確定後あるいは単語認識後音声出力においては、単語確定あるいは単語認識した直後に前記確定あるいは認識した単語に対応した音声を合成すること、文節確定後あるいは認識後の音声出力においては、文節を確定あるいは認識した直後に前記確定あるいは認識した文節に対応した音声を合成すること、文確定あるいは認識後の音声出力においては、文を確定あるいは認識した後に前記確定した文に対応する音声を合成することを特徴とする特許請求範囲第1項記載および第2項記載の音声発生システム。

【請求項6】前記人間の声を基にした音声辞書は音声素片部分、韻律部分の少なくとも1つからなることを特徴とする特許請求範囲第1項記載および第2項記載の音声発生システム。

【請求項7】前記人間の声を基にした音声辞書の前記韻律部分は基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも1つ以上の独立部分を含んでいることを特徴とする特許請求範囲第6項記載の音声発生システム。

【請求項8】前記人間の声を基にした音声辞書の各部分は音声素片、基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも1つ以上の独立部分を含んでいることを特徴とする特許請求範囲第1項記載および第2項記載の音声発生シ

【特許請求の範囲】

【請求項1】入力テキストを解析するテキスト解析手段と解析されたテキスト情報に対して音韻系列を作成する音韻系列作成手段と韻律を作成する韻律作成手段と前記音韻系列と韻律に基づいて合成音声を作成する合成手段を有する音声規則合成により音声を発生させるシステムにおいて

- A : 人間の声を基に音声辞書を作成するステップ
 - B : 前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ
 - C : 前記音声辞書は前記ハード媒体に格納してユーザーに供給され方法及びユーザーの装置が接続可能な回線に結合可能なサーバーを通じてユーザーに供給することから選ばれた少なくとも1つの供給法により供給された前記人間の声を基にした音声辞書をユーザーのシステムの必要に応じて追加インストールおよび削除するステップ
 - D : 前記入力テキストを解析するステップ
 - E : 前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため前記音声辞書の中から任意の音声辞書を選択するステップ
 - F : 前記解析されたテキスト情報に従って前記音声辞書の必用部分を使用し音声を合成するための音声信号データを作成していくステップ
 - G : 前記音声信号データを音声の出力に変換するステップ
- の各ステップを有することを特徴とする音声発生システム。

【請求項2】入力テキストを解析するテキスト解析手段と解析されたテキスト情報に対して音韻系列を作成する音韻系列作成手段と韻律を作成する韻律作成手段と前記音韻系列と韻律に基づいて合成音声を作成する合成手段を有する音声規則合成により音声を発生させるシステムにおいて

- A : 人間の声を基に音声辞書を作成するステップ
 - B : 前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ
 - C : 前記テキストを解析するステップ
 - D : ユーザーが任意の音声を選択するために前記サーバー上の前記音声辞書の中から任意の音声辞書を選択するステップ
 - E : 前記解析されたテキスト情報に従って前記音声辞書の必用部分を使用し音声を合成するための音声信号データを作成していくステップ
 - F : 前記音声信号データを音声の出力に変換するステップ
- の各ステップを有することを特徴とする音声発生システム。

【請求項3】ユーザーは必用に応じてそれぞれの前記人

ステム。

【請求項9】前記可変部分は、声の高さ、発声速度、音量、から選ばれた少なくとも1つ以上であることを特徴とする特許請求範囲第3項記載の音声発生システム。

【請求項10】前記システムにおいて環境音響をデジタル情報に変換し前記環境音響をハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じてバックグラウンド音響としての環境音響を前記合成音声に付加することが可能であることを特徴とする特許請求範囲第1項記載の音声発生システム。

【請求項11】前記システムにおいて環境音響をデジタル情報に変換しユーザーの装置が接続可能な回線に結合されたサーバー上に前記環境音響をデジタル情報に変換した環境音響情報テーブルとして置いておき前記環境音響のデジタル情報をユーザーは必用に応じて前記サーバーに接続し前記バックグラウンド音響としての環境音響を前記合成音声に付加することが可能であることを特徴とする特許請求範囲第2項記載の音声発生システム。

【請求項12】ユーザーの任意な音楽の歌詞に対応したテキストをユーザーの任意の前記音声素片辞書から選ばれた音声素片を接続し音韻系列を作成し前記音韻系列に前記ユーザーの任意な音楽に対応した韻律情報を与えることにより前記音楽に対応した歌声を合成することを特徴とする特許請求範囲第1項記載および特許請求範囲第2項記載の音声発生システム。

【請求項13】前記ユーザーの任意な音楽に対応した歌詞および楽譜から作成した音韻系列および韻律作成のための基情報はハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記ユーザーの任意な音楽に対応した音韻系列および韻律作成のための基情報を使用可能であることを特徴とする特許請求範囲第12項記載の音声発生システム。

【請求項14】前記ユーザーの任意な音楽に対応した歌詞および楽譜から作成した音韻系列および韻律作成のための基情報テーブルはユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記音楽に対応した前記ユーザーの任意な音楽に対応した音韻系列および韻律作成のための基情報を使用可能であることを特徴とする特許請求範囲第12項記載の音声発生システム。

【請求項15】前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換し前記合成音声である歌声に付加可能であることを特徴とする特許請求範囲第12項記載の音声発生システム。

【請求項16】前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換しハード媒体に格納して

ユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した情報を前記合成音声である歌声に付加可能であることを特徴とする特許請求範囲第15項記載の音声発生システム。

【請求項17】前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した楽器演奏情報テーブルをユーザーの装置が接続可能な回線に結合されたサーバー上に置いておき前記ユーザーの任意の音楽に対応した楽器の演奏のデジタル情報に変換した情報をユーザーは必用に応じて前記サーバーに接続し前記ユーザーの任意の音楽に対応した楽器の演奏のデジタル情報に変換した情報を前記合成音声である歌声に付加可能であることを特徴とする特許請求範囲第15項記載の音声発生システム。

【請求項18】前記人の声の基にした音声辞書の基の人に関連した画像および前記入力テキストに関連した画像から選ばれた少なくとも1つの画像を音声発生時にディスプレイ上に表示することを特徴とする特許請求範囲第1項記載および特許請求範囲第2項記載の音声発生システム。

【請求項19】前記人の声を基にした音声辞書の基の人に関連した画像および前記入力テキストに関連した画像から選ばれた少なくとも1つの画像をデジタル情報に変換しハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記画像情報を付加可能であることを特徴とする特許請求範囲第18項記載の音声発生システム。

【請求項20】前記人の声を基にした音声辞書の基の人に関連した画像および前記入力テキストに関連した画像から選ばれた少なくとも1つの画像をデジタル情報に変換し画像情報テーブルとしユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記サーバーに接続し前記画像情報を付加可能であることを特徴とする特許請求範囲第18項記載の音声発生システム。

【請求項21】前記システムにおいて各地方毎に標準化された方言辞書を有することを特徴とする特許請求範囲第1項記載および特許請求範囲第2項記載の音声発生システム。

【請求項22】前記システムにおいて前記方言辞書はテキスト解析部、音声合成部に関する音声辞書から選ばれる少なくとも1つに関連づけられることを特徴とする特許請求範囲第21項記載の音声発生システム。

【請求項23】前記方言辞書をハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能

な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記方言辞書を使用することが可能であることを特徴とする特許請求範囲第21項記載の音声発生システム。

【請求項24】前記方言辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記サーバーに接続し前記方言辞書を使用することが可能であることを特徴とする特許請求範囲第21項記載の音声発生システム。

【請求項25】前記ハード媒体はフロッピーディスク、CD-ROM、MO、PD、DVD、メモリーカードから選ばれた少なくとも1つの媒体であることを特徴とする特許請求範囲第1項記載、特許請求範囲第10項記載、特許請求範囲第13項記載、特許請求範囲第16項記載、特許請求範囲第19項記載および特許請求範囲第23項記載の音声発生システム。

【請求項26】前記回線は一般公衆回線及び専用回線から選ばれた少なくとも1つの回線であることを特徴とする特許請求範囲第1項記載、特許請求範囲第2項記載、特許請求範囲第10項記載、特許請求範囲第11項記載、特許請求範囲第13項記載、特許請求範囲第14項記載、特許請求範囲第16項記載、特許請求範囲第17項記載、特許請求範囲第19項記載、特許請求範囲第20項記載、特許請求範囲第23項記載および特許請求範囲第24項記載の音声発生システム。

【請求項27】前記回線は有線回線および無線回線から選ばれた少なくとも1つの回線であることを特徴とする特許請求範囲第1項記載、特許請求範囲第2項記載、特許請求範囲第10項記載、特許請求範囲第11項記載、特許請求範囲第13項記載、特許請求範囲第14項記載、特許請求範囲第16項記載、特許請求範囲第17項記載、特許請求範囲第19項記載、特許請求範囲第20項記載、特許請求範囲第23項記載および特許請求範囲第24項記載の音声発生システム。

【請求項28】前記有線回線は電気信号および光信号から選ばれた少なくとも1つの有線回線であることを特徴とする特許請求範囲第27項記載の音声発生システム。

【請求項29】前記無線回線は電磁波信号及び光信号から選ばれた少なくとも1つの無線回線であることを特徴とする特許請求範囲第27項記載の音声発生システム。

【請求項30】前記システムにおいてユーザーがテキストの内容に応じて前記テキスト部分、部分で任意の音声を選べることを特徴とする特許請求範囲第1項及び第2項記載の音声発生システム。

【請求項31】前記システムに使用される装置はパソコン、音発生機能付きワープロ、専用装置から選ばれる少なくとも1つの装置であることを特徴とする特許請求範囲第1項記載の音声発生システム。

【請求項32】前記システムに使用される装置はパソコン、音発生機能付きワープロ、専用装置、N C (ネット

ワークコンピューター)、ゲームマシン、STB (セット・トップ・ボックス)から選ばれる少なくとも1つの装置であることを特徴とする特許請求範囲第2項記載の音声発生システム。

【請求項33】前記システムにおける少なくとも1人以上の声を基にした前記音声辞書、前記環境音響をデジタル情報に変換した情報、前記ユーザーの任意な音楽に対応した音韻系列および韻律作成のための基情報、前記ユーザーの任意の音楽に対応した楽器の演奏情報に変換した情報、前記画像情報、前記方言辞書から選ばれた少なくとも1つのデータを格納した前記ハード媒体は売り切りであることを特徴とする特許請求範囲第1項記載、特許請求範囲第10項記載、特許請求範囲第13項記載、特許請求範囲第16項記載、特許請求範囲第19項記載、および特許請求範囲第23項記載の音声発生システム。

【請求項34】前記システムにおける前記ハード媒体は1人以上の人の声を基にした音声辞書、前記環境音響をデジタル情報に変換した情報、前記ユーザーの任意な音楽に対応した音韻系列および韻律作成のための基情報、前記ユーザーの任意の音楽に対応した楽器の演奏情報、前記画像情報、前記方言辞書から選ばれた少なくとも1つのデータを格納しガードをかけ、ユーザーに供給しユーザーの入金を確認後前記ガードを外すことを特徴とする特許請求範囲第1項記載、特許請求範囲第10項記載、特許請求範囲第13項記載、特許請求範囲第16項記載、特許請求範囲第19項記載、および特許請求範囲第23項記載の音声発生システム。

【請求項35】前記システムにおける前記ハード媒体に格納およびユーザーの接続可能な回線に結合されたサーバー上に格納から選ばれた少なくとも1つの格納データは複数の人の声を基にした音声辞書、前記環境音響をデジタル情報に変換した情報、前記ユーザーの任意な音楽に対応した音韻系列および韻律作成のための基情報、前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した情報、前記画像情報、前記方言辞書から選ばれた少なくとも1つのデータであり、それぞれのデータは独立にガードをかけ、ユーザーの入金を確認後それぞれ独立に前記ガードを外すことを特徴とする特許請求範囲第1項記載、特許請求範囲第10項記載、特許請求範囲第13項記載特許請求範囲第16項記載、特許請求範囲第19項記載、および特許請求範囲第23項記載の音声発生システム。

【請求項36】前記システムにおいてユーザーの入金確認により前記サーバー及びサーバー上の辞書、情報の必用部分から選ばれた少なくとも1つのデータに接続可能とすることを特徴とする特許請求範囲第1項記載、特許請求範囲第2項記載、特許請求範囲第10項記載、特許請求範囲第11項記載、特許請求範囲第13項記載、特許請求範囲第14項記載、特許請求範囲第16項記載、

特許請求範囲第17項記載、特許請求範囲第19項記載、特許請求範囲第20項記載、特許請求範囲第23項記載および特許請求範囲第24項記載の音声発生システム。

【請求項37】前記システムにおいてユーザーの入金により前記サーバーに接続し必要音声辞書部分、情報部分をユーザーの装置にコピー可能なことを特徴とする特許請求範囲第1項記載、特許請求範囲第2項記載、特許請求範囲第10項記載、特許請求範囲第11項記載、特許請求範囲第13項記載、特許請求範囲第14項記載、特許請求範囲第16項記載、特許請求範囲第17項記載、特許請求範囲第19項記載、特許請求範囲第20項記載、特許請求範囲第23項記載および特許請求範囲第24項記載の音声発生システム。

【請求項38】前記システムにおいてユーザーの装置にコピーされた前記音声辞書、情報が必要に応じて自動消滅することを特徴とする特許請求範囲第1項記載、特許請求範囲第10項記載、特許請求範囲第13項記載、特許請求範囲第16項記載、特許請求範囲第19項記載、および特許請求範囲第23項記載の音声発生システム。

【請求項39】前記システムにおいてユーザーが前記サーバーに接続し前記サーバー上の音声辞書、情報の使用量に応じて料金を決定し、前記料金をクレジットカード、銀行引き落とし、請求書発行による入金から選べた少なくとも1つの決済法をとることを特徴とする特許請求範囲第1項記載、特許請求範囲第2項記載、特許請求範囲第10項記載、特許請求範囲第11項記載、特許請求範囲第13項記載、特許請求範囲第14項記載、特許請求範囲第16項記載、特許請求範囲第17項記載、特許請求範囲第19項記載、特許請求範囲第20項記載、特許請求範囲第23項記載および特許請求範囲第24項記載の音声発生システム。

【請求項40】前記システムによって作成した音声をオーディオカセット、MD、DCCなどの録音媒体に記録することを特徴とする特許請求範囲第1項記載および特許請求範囲第2項記載の音声発生システム。

【請求項41】前記システムにおいてパソコン通信におけるチャットのテキストを音声に変換することを特徴とする特許請求範囲第1項記載の音声発生システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明はパソコン、ワープロ、ゲーム機、専用機、NC（ネットワークコンピューター）、STB（セット・トップ・ボックス）などを利用する際の合成音声の発生に係わり、特にユーザーが任意でかつ多様な合成音声の発生を選ぶことが可能な手段を実現するシステムに関する。

【0002】

【従来の技術】従来の音声合成に関しては、多種多様の装置及び方法が提案されておりまた構内放送、ゲーム情

報サービス等の様々な分野で広く用いられている。音声は言語として意味をなす最小単位である音素から構成されている。この音素は波形として表現処理される。その波形の一部あるいは全部が音声の合成に使用される音声素片となる。この音声素片を合成単位として任意の音声を作る合成方式では、各音声素片を一定の規則に従って結合し、合成音声を発生させている。こういった方法では、従来基本周波数を変えて男性音声、女性音声を区別して作り出す方法、時間軸を伸縮させて発生速度を変える方法などが提案されており、こういった声質を変化させる音声合成についても実際に商品化されている。実際の商品としてはドキュメントトナー、TTSなどの商品などがありテキストファイルを読み上げる音声合成の商品として実際商品化されておりパソコン上で音声合成を実現している。

【0003】自然な言語を実現するために各音声素片の接合の工夫も種々の工夫が提案されており人が実際に発声するように自然な発声が実現されている。また文書の意味を解析しその結果を基にして、重要な部分を強調するためにその部分の語調を強めたりする音声合成方法など各種の方法が提案されている。しかしながら現行の方法では合成音声の声質は基本周波数を変更したりするものに限られており多種多様の声質の要求に満足できるものでは無かった。

【0004】

【発明が解決しようとする課題】しかしながら人が聞いて違和感のない自然な音声はいつも聞いている人の声をベースにしたものの方が自然である。またいつも同質の声ではメリハリがつかないため、或いは状況などによって聞きたい声質は異なるため、色々な人の声をベースにした音声を状況によって使い分け種々な音声を発声させることができていた。

【0005】人にはそれぞれ声質、基本周波数、アクセント、イントネーション、ポーズの取り方、言葉の一部を伸ばしたりする時間的特徴、固有の方言の特徴、感情により音の高低、音量の変化などの特徴が異なる。こういった人により異なる特徴的要素を加えることが可能となればより臨場感のある音声発声がユーザーの状況に合わせ自由に選べるはずである。

【0006】各地方の方言による音声合成を任意の人の声をベースにしたもの聞きたいなどということも望まれていた。また任意の合成音声を雰囲気或いは状況に応じた任意のバックグラウンド音響の中で聞くということも望まれていた。同様に任意の声質で歌を歌わせることも望まれている。前記ユーザーが任意に選択した声質に関連した画像をディスプレイ上に表示することにより更に臨場感を増すことが可能となる。

【0007】特に最近のパソコンの性能のアップにおいては、目を見張るものがあるが、前記音声合成をパソコン上で実現する事は、非常に有意義なことである。また

パソコンだけではなくゲーム機において本技術が実現されれば非常に臨場感のあるゲームが実現する。また新しい概念であるNC(ネットワークコンピュータ)などでも新しいアプリケーションが広がる可能性がある。

【0008】

【課題を解決するための手段】前記目的を達成するために、本発明においては次に示す方法をとる。すなわち予めテキストファイルである情報、キーボードなどの入力装置によって入力されたテキストに変換された情報、OCR(オプティカル・キャラクター・リーダー)によって活字からのテキストに変換された情報の少なくとも1つの情報のテキスト入力を解析するテキスト解析手段と解析されたテキストデータに対して音韻系列を作成する音韻系列作成手段と韻律を作成する韻律作成手段と前記音韻系列と韻律に基づいて合成音声を作成する合成手段を有する音声規則合成を使用して音声を発生させるシステムにおいてユーザーの任意の音声を発生させるシステムであり、そのステップはユーザーの装置がパソコン、音発生機能付きワープロ、専用装置から選ばれた少なくとも1つの装置である場合は

- A : 人間の声を基に音声辞書を作成するステップ
- B : 前記音声辞書をハード媒体に格納してユーザーに供給すること及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つのユーザーへの音声辞書供給ステップ
- C : 前記音声辞書は前記ハード媒体に格納してユーザーに供給され方法及びユーザーの装置が接続可能な回線に結合可能なサーバーを通じてユーザーに供給することから選ばれた少なくとも1つの供給法により供給された前記人間の声を基にした音声辞書をユーザーのシステムの必要に応じて追加インストールおよび削除するステップ
- D : 前記入力されたテキストを必用に応じて解析するステップ

E : 前記ユーザーのシステムにおいてはユーザーが任意の音声を選択するため1以上の音声辞書を有している前記音声辞書の中から任意の音声辞書を選択するステップ

F : 前記テキストを必用に応じて解析された情報に従つて前記音声辞書の必用部分を使用し音声を合成するためのデータを作成していくステップ

G : 前記音声を合成するためのデータを音声の出力に変換する手段を有するステップ
の各ステップからなるシステムとする。

【0009】またユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、NC(ネットワークコンピュータ)、ゲームマシン、STB(セット・トップ・ボックス)から選ばれる少なくとも1つの装置であり、JAVAなどの言語を使用しユーザーのシステムにおいては最小限のソフトのみを持ち、回線に接続可能なサーバー上にソフトおよび各種辞書が格納されている場合には

A : 人間の声を基に音声辞書を作成するステップ

B : 前記音声辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に設置するステップ

C : 前記入力されたテキストを必用に応じて解析するステップ

D : ユーザーが任意の音声を選択できるようにユーザーの装置が接続可能な回線に結合されたサーバー上の音声辞書の必要部分を選択するステップ

E : 前記テキストを必用に応じて解析された情報に従つて前記音声辞書の必用部分を使用し音声を合成するためのデジタルデータを作成していくステップ

F : 前記音声を合成するためのデータを音声の出力に変換する手段を有するステップ
の各ステップからなるシステムとすればよい。

【0010】そして前記システムにおいてはユーザーは必用に応じてそれぞれの前記人間の声を基にした音声辞書をユーザーの任意の組み合わせで選べる構造とする。前記人間の声を基にした音声辞書から選択した音声辞書の各要素の中から必用に応じて前記音声の可変可能部分の少なくとも1つをユーザー好みで可変させることを可能とする。この可変部分、声の高さ、発声速度、音量、から選ばれた少なくとも1つ以上である。

【0011】本発明によるシステムにおいては入力がキーボード入力、OCRを利用した活字文字の認識、手書き文字の認識入力から選ばれる少なくとも1つの方法の入力法であり、1文字入力の後の音声出力、単語確定後の音声出力、文節確定後の音声出力、文確定後の音声出力、1文字認識の後音声出力、単語認識後の音声出力、文節認識後の音声出力、文認識後の音声出力の少なくとも1つをユーザーの必用に応じて選択できることが可能である。前記1文字入力の後音声出力の場合にはアルファベット、ひらがな、及びカタカナから選ばれる少なくとも1つの方法で入力された1つの日本語文字を認識した直後に前記認識した文字に対応した音声を発生させること、単語確定あるいは単語認識後音声出力においては、単語確定あるいは単語認識した直後に前記確定あるいは認識した単語に対応した音声を発生させること、文節確定あるいは文節認識後音声出力においては、文節を確定あるいは認識した直後に前記確定あるいは認識した文節に対応した音声を発生させること、文確定あるいは認識した文節の音声出力においては、文を確定あるいは認識した後に前記確定した文に対応する音声を発生させることができあるシステムとする。

【0012】前記人間の声を基にした音声辞書は音声素片部分、韻律部分から構成され、必用に応じてそれぞれの部分に分けて格納しておく。

【0013】前記人間の声を基にした音声辞書の前記韻律部分は基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも

1つ以上の独立部分を含んでおりそれぞれの部分に分けて格納しておく。

【0014】あるいは前記人間の声を基にした音声辞書の各部分は音声素片、基本周波数、アクセント、イントネーション、時間的特徴、音量の特徴、方言的特徴、ポーズの取り方の特徴、感情による発声の特徴から選ばれる少なくとも1つ以上の独立部分を含ませてもよい。

【0015】前記手法により作成した任意の声質の合成音声を任意の特定の場所あるいは音楽などのバックグラウンド音響を付加するために以下の手法をとることも可能である：すなわちユーザーの装置がパソコン、音発生機能付きワープロ、専用装置から選ばれた少なくとも1つの装置である場合は前記システムにおいて環境音響をデジタル情報に変換し前記環境音響をハード媒体に格納してユーザーに供給され方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じてバックグラウンド音響としての環境音響を前記合成音声に付加すればよい。またユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、N C、ゲームマシン、S T B（セット・トップ・ボックス）から選ばれた少なくとも1つの装置でありユーザーの装置が接続可能な回線に結合されたサーバー上にソフトおよび各種辞書が格納されている場合には環境音響をデジタル情報に変換し環境音響テーブルとし回線に結合されたサーバー上に置いておき前記環境音響のデジタル情報をユーザーは必用に応じて前記サーバーに接続しバックグラウンド音響としての環境音響を前記合成音声に付加すればよい。

【0016】ユーザーの任意の声質で歌声を発声させるためには以下の方法をとればよい。すなわち前記音声素片辞書から選ばれ前記ユーザーの任意な音楽に対応したテキストをユーザーの任意の前記音声素片辞書から選ばれた音声素片を接続し音韻系列とし前記ユーザーの任意な音楽に対応した韻律を韻律情報に変換し前記ユーザーの任意な音楽に対応した韻律情報を与えることにより前記音楽に対応した歌声を合成すればよい。ここでユーザーの装置がパソコン、音発生機能付きワープロ、専用装置から選ばれた少なくとも1つの装置である場合は前記ユーザーの任意な音楽に対応した韻律をデジタル情報に変換した情報はハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記ユーザーの任意な音楽に対応した韻律をデジタル情報に変換した情報を使用することが可能となる。またユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、N C、ゲームマシン、S T B（セット・トップ・ボックス）から選ばれる少なくとも1つの装置であり回線に接続可能なサーバー上にソフトおよび各種辞書が格納され

ている場合には前記ユーザーの任意な音楽に対応した韻律をデジタル情報に変換した情報は回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記音楽に対応した韻律をデジタル情報に変換した情報を使用するようすればよい。

【0017】上記歌声に楽器の演奏を付加することも可能である。すなわち前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換し前記音声合成である歌声に付加すればよい。そしてユーザーの装置がパソコン、音発生機能付きワープロ、専用装置から選ばれた少なくとも1つの装置である場合は前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換しハード媒体に格納してユーザーに供給される方法及び回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは自分の装置に前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した情報をインストールし必用に応じて前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した情報を前記合成音声である歌声に付加すればよい。ユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、N C、ゲームマシン、S T B（セット・トップ・ボックス）から選ばれる少なくとも1つの装置でありユーザーの装置が接続可能な回線に結合されたサーバー上にソフトおよび各種辞書が格納されている場合には前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換しユーザーの装置が接続可能な回線に結合されたサーバー上に置いておき前記ユーザーの任意の音楽に対応した楽器の演奏のデジタル情報に変換した情報をユーザーは必用に応じて前記サーバーに接続し前記ユーザーの任意の音楽に対応した楽器の演奏のデジタル情報に変換した情報を前記合成音声である歌声に付加する。

【0018】前記人の声を基にした音声辞書の基にした人に関連した画像および前記入力テキストに関連した画像から選ばれた少なくとも1つの画像をディスプレイ上に表示することも可能であり、ユーザーの装置がパソコン、音発生機能付きワープロ、専用装置から選ばれた少なくとも1つの装置である場合は前記人の声を基にした音声辞書の基の人に関連した画像および前記入力テキストに関連した画像から選ばれた少なくとも1つの画像をデジタル情報に変換しハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記人の声を基にした音声辞書の基の人に関連した画像および前記入力テキストに関連した画像から選ばれた少なくとも1つの画像をデジタル情報に変換した情報を付加すればよい。またユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、N C、ゲームマシン、S T B（セット・トップ・ボックス）から選ばれる少なくとも

1つの装置であり回線に接続可能なサーバー上にソフトおよび各種辞書が格納されている場合には前記人の声を基にした音声辞書の基の人に関連した画像および前記入力テキストに関連した画像から選ばれた少なくとも1つの画像をデジタル情報に変換し画像情報テーブルとしてユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記サーバーに接続し前記人の声を基にした音声辞書の基の人に関連した画像デジタル情報に変換した情報を付加すればよい。

【0019】次に音声発生において方言的な特徴も1つの興味深い要素である。これを実現するには各地方毎に標準化された方言の辞書を設ければよい。そして前記各地方毎に標準化された方言の辞書はテキスト解析部、音声合成部に関する音声辞書から選ばれる少なくとも1つに関連づけられるようとする。ユーザーの装置がパソコン、音発生機能付きワープロ、専用装置から選ばれた少なくとも1つの装置である場合は前記各地方毎に標準化された方言の辞書をハード媒体に格納してユーザーに供給される方法及びユーザーの装置が接続可能な回線を通じてユーザーに供給することから選ばれた少なくとも1つの供給法によりユーザーは必用に応じて前記各地方毎に標準化された方言の辞書を使用すればよい。ユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、NC、ゲームマシンSTB（セット・トップ・ボックス）から選ばれる少なくとも1つの装置であり回線に接続可能なサーバー上にソフトおよび各種辞書が格納されている場合には前記各地方毎に標準化された方言の辞書をユーザーの装置が接続可能な回線に結合されたサーバー上に置いておきユーザーは必用に応じて前記サーバーに接続し前記各地方毎に標準化された方言の辞書使用すればよい。

【0020】各種辞書、各情報テーブルをユーザーに供給するためのハード媒体はフロッピーディスク、CD-ROM、MO、PD、DVD、メモリーカードからその目的によって最適なものを選択すればよい。

【0021】同様に回線を通じてユーザーに供給する場合および回線に接続可能なサーバー上にソフト、各種辞書、各情報テーブルが格納されている場合の回線は一般公衆回線及び専用回線から選ばれた少なくとも1つの回線であり、前記回線は必用に応じて有線回線および無線回線からあるいは組み合わせて最適なものを選択すればよい。前記有線回線は電気信号および光信号から選ばれた少なくとも1つの有線回線であり、前記無線回線は電磁波信号及び光信号から選ばれた少なくとも1つの無線回線であればよい。

【0022】本発明のシステムにおいて文章を元に音声を発声させる場合は文章の各部分部分で独立に任意の声質の音声を選択できる構造とする。

【0023】本発明のシステムによるユーザーへのハード媒体の供給法は前記システムにおける少なくとも1人

以上人の声を基にした前記音声辞書、前記環境音響をデジタル情報に変換した情報、前記ユーザーの任意な音楽に対応した韻律のデジタル情報に変換した情報、前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した情報、前記人の声を基にした音声辞書の基の人に関連した画像をデジタル情報に変換した情報、前記各地方毎に標準化された方言の辞書から選ばれた少なくとも1つの情報を格納した前記ハード媒体は売り切りとする。

【0024】あるいは前記システムにおける前記ハード媒体は1人以上の人の声を基にした音声辞書、前記環境音響をデジタル情報に変換した情報、前記ユーザーの任意な音楽に対応した韻律のデジタル情報に変換した情報、前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した情報、前記人の声を基にした音声辞書の基の人に関連した画像をデジタル情報に変換した情報、前記各地方毎に標準化された方言の辞書から選ばれた少なくとも1つの情報を格納しガードをかけ、ユーザーの入金を確認後前記ガードを外すシステムとしてもよい。

【0025】または前記システムにおける前記ハード媒体は複数の人の声を基にした音声辞書、前記環境音響をデジタル情報に変換した情報、前記ユーザーの任意な音楽に対応した韻律のデジタル情報に変換した情報、前記ユーザーの任意の音楽に対応した楽器の演奏をデジタル情報に変換した情報、前記人の声を基にした音声辞書の基の人に関連した画像をデジタル情報に変換した情報、前記各地方毎に標準化された方言の辞書から選ばれた少なくとも1つの情報を格納しそれぞれの前記人の声を基にした音声辞書独立にガードをかけ、ユーザーの入金を確認後それぞれ独立に前記ガードを外すシステムでもよく必用によって最適なものを選択すればよい。

【0026】回線を通じてユーザーに供給する方法あるいは、ユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、NC、ゲームマシンSTB（セット・トップ・ボックス）から選ばれる少なくとも1つの装置であり回線に接続可能なサーバー上にソフトおよび各種辞書が格納されている場合にはユーザーの入金により前記サーバーに接続可能とすればよい。

【0027】回線を通じてユーザーに供給する方法においてはユーザーの入金により前記サーバーに接続し必要音声辞書部分をユーザーの装置にコピー可能とすればよい。そしてこの場合ユーザーの装置にコピーされた前記音声辞書が必要に応じて自動消滅する構造でもよい。前記構造を必用に応じて自由に組み合わせればよい。

【0028】またユーザーのシステムがパソコン、音発生機能付きワープロ、専用装置、NC、ゲームマシン、STB（セット・トップ・ボックス）から選ばれる少なくとも1つの装置であり回線に接続可能なサーバー上にソフトおよび各種辞書が格納されている場合にはユーザ

ーが前記サーバーに接続し前記サーバー上の音声辞書の使用量に応じて料金を決定し、前記料金をクレジットカード、銀行引き落とし、請求書発行による入金から選ばれた少なくとも1つの決済法をとればよい。

【0029】上述したシステムで発生した音声をユーザーの公用に応じてオーディオカセット、MD、DCCなどの録音媒体に記録してもよい。また前記システムをパソコン通信におけるチャットで使用してもよい。

【0030】

【作用】前述のシステムによれば人の声をベースにした各特徴を独立に有する音声辞書および各情報を各種ユーザーに容易に供給できる。また各ユーザーにおいては、雰囲気、状況などにより任意の声質、語調などを任意の自由に音声を選ぶことが可能となる。各地方の方々言、または人により異なる方言的特徴なども加味した任意の音声で実現できる。また各音声辞書と同様にバックグラウンド音響もユーザーに容易に供給できるのでユーザーの任意のバックグラウンド音響で任意の音声を聞くことが可能となる。またユーザーは任意の声質で任意の歌を歌わせることができ、かつ任意の楽器の伴奏も付加することが可能となる。任意の声質の基の人に関連した画像をディスプレイ上に表示することができるのにより臨場感が出る。この合成音声を録音媒体に記録することにより携帯用のオーディオカセットテープにより繰り返して再生させることができとなり学習などにも役に立つ。

【0031】

【実施例】図1に本発明の実施例の全体のシステムの構成の概要を示す。ここで1は音声提供者であり、まず予めモデル化された文章を音声提供者に一定のテンポで朗読して貰う。朗読時に一定のテンポで朗読してもらうために音声提供者にはヘッドフォンを付けてもらいこのヘッドフォンよりモデルの文書と朗読のリズムの指示を流す。この音声提供者の朗読の結果を2の音声記録装置に記録する。次にこの朗読の結果を元に音声辞書を作成していく。そしてその結果を3のコンピュータに格納する。この辞書を4に示すようにCD-ROMに格納する。

【0032】そしてこのCD-ROMをユーザに供給し、ユーザーのパソコン5にCD-ROMに格納されている辞書をインストールする。もちろん音声辞書のユーザーへの供給法については、CD-ROMである公用ではなくFD、DVD、PD、メモリーカードなどでもよい。また図の5のサーバーに音声辞書、情報テーブルを格納しておきユーザーが一般電話回線を経由してサーバーに接続して任意の音声辞書を自分のパソコンにインストールしてもよい。

【0034】図2にユーザーが音声合成していく様子を概略フローチャート示す。まずユーザーは原稿をスキナーを用いてパソコンに取り込む。次にこの情報をテキス

トファイルに変換する。このテキストファイルに対してテキスト解析を行う。このテキスト解析の結果に基き21のユーザーの任意の音声素片辞書を参照し前記テキストを解析した結果にそって各音声素片を接続していき接続部を補完し音韻系列を作成する。この音韻系列と22の韻律辞書を参照しユーザーの任意の韻律を選択して作成した韻律を合成し音声合成データを作成する。

【0035】次に環境音響を付与する。ここでは海の波の音をデジタル情報としたものを音声と同様の方法すなわちCD-ROMに格納してユーザーに供給しユーザーは自分のパソコンの環境音響情報テーブル23にインストールする。前述の方法で作成した合成音声に環境音響情報テーブル23を参照してユーザーの任意の海の波の音を付加する。

【0036】前述の処理により作成したデータをパソコンの音源ボードを使用し音声出力に変換しスピーカーより音として出力する。

【0037】本実施例ではユーザーの装置はパソコンであったがパソコンに限定されるものではなく例えば音発生機能付きワープロ、音発生機能付きゲーム機、あるいは専用機などでもよい。

【0038】更に現在提案されている安価なコンピューターNC(ネットワークコンピュータ)、STB(セット・トップ・ボックス)などを使用してもよい。この場合公用なソフト、音声辞書、情報テーブルは一般公衆電話回線などに結合されたサーバー上に設置しておきユーザーが公用な時ののみ前述のサーバーに接続し使用すればよい。

【0039】次に各部についてさらに詳細に述べる。図3は音声辞書の構造を示す。ここでそれぞれ301は音声素片辞書であり302は韻律辞書である。さらにこの韻律辞書部分は303の基本周波数部分、304の発音される個々の語が持つ強さまたは高さの配置であるアクセント部分、305の話をするととき話し手の上がり下がりの調子であるイントネーション部分、306の時間的特徴部分、307の声の大きさを表す音量の特徴、308の方言的特徴、309のポーズの取り方の特徴、310の感情による発声の特徴の独立部分を備えている。

【0040】図4に音声素片辞書の作成のフローチャートを示した。音声提供者に予め決められた文章を朗読してもらう。この際音声提供者にはヘッドホンを装着してもらい、このヘッドフォンを通じて文書の朗読のタイミングを指示する。音声提供者はこの指示に基づいて前述の予め決められた文章を朗読していく。この朗読した音声を図4の401に示すようにアナログ録音する。次に402においてアナログの音声をデジタル変換していく。このデジタル変換された音声データを403に示すように時間分割処理を行う。次に404に示すように予め決められた文書の音素と音声提供者の音素の対応付けを行う。この対応付けを行った結果を元に各音素にラベル

グを行う。この結果を音素ラベリングテーブル408に格納する。音素ラベリングテーブルおよび予め作成してある音声素片切り出しテーブル409に従い406で音声素片を切り出す。切り出された音声素片は407で示された格納処理部において辞書化処理を行い音声素片辞書410に格納する。

【0041】韻律辞書も音声素片辞書と同様に予め決められた文章を音声提供者に読んでもらう。但し音声素片辞書作成の時と異なるのは読む際のタイミングの指示を音声提供者に行わない。また興奮した場合、悲しい場合、怒った場合等の感情の起伏を予めきめられた文書の必用部分に添付しておく。そして韻律辞書の場合には標準の韻律を定めておき各音声提供者の辞書は標準との差の形で作成する。この時予め標準として定めたの男性の声と女性の声の韻律辞書を作成しておく。但し標準の韻律辞書の感情部分は抑揚のない状態での基本周波数、音量、発話速度を格納しておく。

【0042】図5の(a)に標準の音声波のアウトライン、(b)に音声提供者Aの音声波のアウトラインを示す。基本周波数の採取については音声提供者に抑揚のない平板な状態で読んで貰う。基本周波数については標準の基本周波数に対して(b)の音声提供者Aの音声波をそれぞれ予め設定された語についての基本周波数を比較し、その結果を音声辞書の韻律部の基本周波数の部分に格納する。

【0043】次にアクセントについては音声提供者にいつも話している状態で読んでもらう。予めそれぞれ設定された各単語の各部分のピーク部分の位置、高さおよび基本周波数を標準の音声波および音声提供者Aの音声波を比較し標準し、この結果の差を音声辞書の韻律部のアクセントの部分に格納する。ここでピークの音声の位置は図5においては標準のものでは502であり、音声提供者Aにおいては512である。また高さ（音声の強さ大きさ）は標準のものでは501であり、音声提供者Aにおいては511である。

【0044】イントネーションについては予め決められた文を音声提供者にいつも話している状態で読んでもらう。そして基本周波数の変化を標準のものと比較しその結果の差を音声辞書の韻律部のイントネーションの部分に格納する。図6にその様子を示した。ここで図5の(a)の標準の基本周波数の変化を図6(a)に示すようにFとして縦軸にとった。同様に図5の(b)の音声提供者Aの基本周波数の変化を図6(b)に示すようにFとして縦軸にとった。

【0045】時間的特徴については予め決められた文章中の各音および代表的単語の継続時間時間、てにおはおよび接続詞の最後の音の継続時間そして文章全体を読む早さを比較する。前記と同様に標準のものと比較しその結果の差を音声辞書の韻律部の時間的特徴の部分に格納する。

【0046】方言的特徴については音声提供者が通常の話す際に方言が出る場合には出身地方の標準の方言辞書との比較を行いその結果の差を音声辞書の韻律部の方言部分に格納する。

【0047】ポーズの特徴については図5に示すポーズ部の継続時間、標準では504、音声提供者Aにおいては514部分を比較する。そしてその結果の差を音声辞書の韻律部のポーズ部分に格納する。

【0048】感情による特徴については音声提供者に前述のように興奮した場合、悲しい場合、怒った場合等の感情の起伏を予め決められた文書の必用部分に添付しておく。そしてこの感情をこめてこの文章を読んで貰う。そしてこの部分を標準のものと比較する。この時比較の対象は文章を読む早さ、基本周波数、音量（音の強さ）を比較する。そしてこの結果を音声辞書の韻律部の感情部分に格納する。

【0049】任意の人の声で音声合成したものに任意のバックグラウンド音響を附加するとさらに臨場感が増す。このために環境音響についても環境音響情報テーブルを作成する。前述したように海の波の音、駅構内の音、高速道路の脇の車の風きり音を図7に示したようにアナログ録音したものを合成音声に附加できる形にデジタル変換し環境音響情報テーブル71に格納すればよい。

【0050】任意の人の音声に関連した画像あるいは入力テキストに関連した画像、又はその双方を音声発生時にディスプレイ上に表示することも効果的である。これを実現するためには、画像情報テーブルを必用とする。これも前述した環境情報テーブルと同様に作成すればよい。例えば任意の人の声の基の声の音声提供者がアニメーションの主人公の声優であれば、そのアニメーションの主人公の話をしている状態の画像を取り込み画像情報テーブルに格納すればよい。また音声提供者自身をスチルカメラでとり静止画像として取り込んでもよいし、またビデオカメラで話をしている所を取り込んでもよく、これを図8に示すように音声発生データに付加できる形に変換し画像情報テーブルに格納すればよい。入力テキストに関連した画像も同様に、その入力テキストが列車の旅の場面であれば列車をスチルカメラあるいはビデオカメラで撮影しておき音声発生データに付加できる形に変換し画像情報テーブルに格納すればよい。

【0051】各地方毎に標準化された方言辞書は韻律だけでなく一部の単語も置き換わる場合があるので置き換え用の単語辞書も必用となる。そこで方言辞書の構成は図9に示すように単語置き換え部と韻律部から構成すればよい。単語置き換え部は標準語の単語に対応する各地方の独特の単語を採取しその地方の単語のテキストと語音の標準語との差を方言辞書の単置き換え部に格納する。同様に例えば発話速度が早かったり語尾を上げるといった全体の韻律も採取し標準語の韻律との差を方言辞

19

書の韻律部に格納する。

【0052】図10のブロック図においてユーザーのシステムについてさらに詳しく説明する。まず紙面上の活字をスキャナーで読み込みシステムに取り込む。これをテキストに変換する。次にこのテキストを標準辞書1001を参照して1011のテキスト解析部で解析する。ここで1010はユーザーシステムのシステム制御部である。この結果を基に任意の人の声を基にした音声辞書の音声素片部分1003(1)を参照して音韻系列生成部1013により音韻系列を作成する。同様に標準の韻律辞書1002および任意の人の声を基にした音声辞書の韻律生成部1003(2)を参照し韻律を生成する。そしてこの生成した音韻系列および韻律から音声合成データを作成する。

【0053】ここでユーザーが方言を指定した場合はシステム制御部1010より指令を出し、まず標準化された方言辞書1004の単語置き換え部を参照して標準語に対応する単語を特定地方の単語テキストに置き換える。次に韻律生成部においても標準化された方言辞書1004の韻律部分および単語置き換え部の語調を参照し前述の単語に対する韻律を生成すると共に、文全体の韻律も標準辞書1001の韻律部分および任意の人の声を基にした音声辞書1003(2)の方言部分を参照して韻律を生成させる。

【0054】またユーザーが例えば怒った感情を付加したい場合には、韻律の一部を可変すればよい。システム制御部1010より指令を出し標準辞書1001の韻律部分および任意の人の声を基にした音声辞書1003(2)の感情部分を参照して韻律修正部1015により修正した韻律を生成させる。さらにユーザーが怒った感情を更に強調したい場合には基本周波数を5%増加し、発話速度を10%早くし、音量を30%増加させる。

【0055】ユーザーがバックグラウンド音響として前述の合成音声に例えば駅構内の音を付加したい場合にはシステム制御部1010より指令を出し環境音響情報テーブル1005を参照して駅の構内の音を環境音響付加部1017において環境音響を付加し音声データを作成する。

【0056】同様にユーザーが任意の人の声を基にした人に関連した画像をディスプレイ上に表示したい場合には付加したい場合にはシステム制御部より指令を出し画像情報テーブル1006を参照して任意の人の声を基にした人に関連した画像を選択し、前述の音声データに画像データを付加し、画像付き音声データとする。

【0057】次にこれを実音声変換部1019において音声データを実音声に変換しスピーカー1007により実音声化すると共に画像処理部1020により画像処理し任意の人の声を基にした人に関連した画像をディスプレイ上に表示する。

【0058】さらに各テキストの部分で声質の異なる場

(11) 20

合の音声発生の手順を説明していく。ここではCD-ROMの形で関西地方の標準化された方言辞書、A氏、B氏の声を基にした音声辞書をユーザーのシステムのハードディスクにインストールした。A氏の声を基にした音声素片辞書は1003(1)のA部、韻律辞書は1004(2)のA部に、B氏の声を基にした音声素片辞書は1003(1)B部に、韻律辞書は1004(2)のB部に格納されている。同様に環境音響情報テーブルには駅構内の音を、画像情報テーブルにはそれぞれA氏、B氏の話をしている動画をインストールしておく。

【0059】図11にこの場合の時間的流れと出力したい音声を示した。ここで時間の流れは左から右であり、1101は標準語のA氏の声、1102は関西弁のB氏の声、そして1103はA氏の声質にB氏の語調を選択した。この指令は図10のシステム制御部1010により、各部へ送られ必用なテキストを解析し1101部分の合成は音声辞書の音声素片辞書部分の1003(1)のA部を参照して音韻系列を生成し、音声辞書の韻律部分の1003(2)のA部を参照して韻律を生成する。

20 20 次に1102部分の合成はテキスト解析した結果を方言辞書1004を参照して単語置き換えを行う。次にこの結果を基に音声辞書の音声素片辞書部分の103(1)B部を参照して音韻系列を生成し、標準方言辞書1004、標準辞書1001の韻律部分および音声辞書の韻律部分の1004(2)B部を参照して韻律を生成する。

【0060】次に1103の部分の音声合成について説明する。この部分は標準語でB氏の声質A氏の語調で話している状態である。この場合にはまず音声辞書の音声素片辞書部分の1003(1)B部を参照して音韻系列を生成し、音声辞書の韻律部分の1003(2)A部を参照して韻律を生成する。そして前述の音韻系列および韻律からそれぞれ1101、1102、1103の部分の合成音声データを生成する。

【0061】前述の合成音声のデータに駅構内の音を付加する。これは図9の環境音響情報テーブルを参照して駅構内の音を付加し音声発生データとする。

【0062】次にA氏、B氏の画像を付加していく。それぞれ1101はA氏の動画、1102および1103はB氏動画を付加すればよいかから、図10の画像情報テーブル1006を参照して前述の音声発生データに画像データを付加し画像データ付き音声発生データとする。

【0063】この画像データ付き音声発生データ処理し音声はスピーカーから、画像はディスプレイに表示する。もちろん入力テキストに関連した画像の場合も同様である。

【0064】次に歌声を発生させる方法について述べる。通常歌についての情報は楽譜に記載されている。楽譜の情報には歌詞、音符、音楽記号があり、歌詞にはテキスト情報、音符には音の高さ(周波数)と長さ、音楽記号には強弱記号が含まれる。まず前述の楽譜情報から

21

歌声を合成するための情報を作成する必用がある。

【0065】楽譜をスキャナーあるいは手入力でコンピュータに取り込む。そして歌詞の情報をテキストに変換する。同様に音符情報、および音楽記号情報から韻律を作成するための基情報を作成する。

【0066】前述の音楽に対応する楽器の演奏も付加すればさらに臨場感が出る。これを実現するために前述の音楽の楽器演奏を録音し歌声に付加できる形にデジタル変換しておけばよい。この楽器演奏の情報には歌声と同期するためにタイミング信号を挿入しておく。

【0067】前述の方法で作成した情報をCD-ROMに格納しユーザーに供給する。ユーザーはこの情報を自分のパソコンのハードディスクに格納する。図12にユーザーが歌聲音を発生させるまでのフローチャートを示した。まず1201にある歌情報からユーザーの任意の歌情報テーブルを参照し1211でテキスト解析を行う。次にこの結果をもとに音声素片辞書1202を参照してユーザーの任意の声質を選び1212で音韻系列を作成する。1212では歌情報テーブル1201の韻律作成のための基情報部分を参照して韻律を作成し、両者を用いて音声合成データを1214で作成する。つぎに前述の歌声データに前述の音楽に対応した任意の楽器演奏情報を1203の楽器演奏情報テーブルを参照して附加する。

【0068】次に音声発生ソフトおよび各辞書のユーザーへの供給方法および課金の方法についてのシステムを説明する。前述のシステムにおいてはユーザーのシステムがパソコンの場合を説明したが音声発生ソフトおよび各辞書が回線に結合されたサーバー上に常駐させ、必用に応じてユーザーの装置から前記サーバーに接続して前記音声発生ソフト、各辞書、各情報テーブルを使用するNC(ネットワークコンピューター)的使用の場合と異なるため分けて説明する。まずユーザーのシステムがパソコンの場合には、大きく分けて2つの方法があり、1つはハード媒体で供給する方法であり、もう1つは回線で供給する方法である。

【0069】前者においてはまず音声発生ソフトおよび音声辞書、環境音響情報テーブル、画像情報テーブルの各辞書、各情報テーブルを一切ガードをかけずにCD-ROMなどの媒体に格納する。そしてこのCD-ROM媒体をユーザーに売り切りで販売する方法がある。またこの音声発生ソフトおよび各辞書それぞれに異なった暗証番号でガードが外れるようにしてインストールする。そしてこの場合CD-ROM媒体を低価格で販売するか、あるいは無料で配布しユーザーは必用部分のみの代金を払い、供給者はこの代金の入金を確認したらユーザーに必用部分の暗証番号を発行する。この暗証番号発行方法は公衆回線を通じてパソコン通信で配布すればよい。もちろん郵送、電話で告げるなどの方法でもよい。ユーザーはこの暗証番号を利用してCD-ROM媒体の

22

必用部分のガードをはずしユーザーのパソコンシステムのハードディスクにインストールする。

【0070】公衆電話回線を通じて供給してもよい。音声発生ソフトおよび各辞書を公衆電話回線に結合したサーバー上に常駐させておく。ユーザーは必用とするソフト、辞書、情報テーブルを前述の回線を通じて入手すればよい。この場合代金の回収は前述のCD-ROMの場合と同様に音声発生ソフトおよび各辞書それぞれに異なった暗証番号でガードをかけておきサーバー上に常駐させておけばよい。そしてユーザーは自由に前記音声発生ソフト、各辞書、各情報テーブルの必用部分を前記回線を通じて自分の装置に取り込む。前述のCD-ROMの場合と同様な方法でガードをはずしユーザーは自分のシステムに必用部分をインストールすればよい。

【0071】ユーザーのシステムに無償あるいは有償でCD-ROM媒体あるいは公衆電話回線を通じてインストールされた音声発生ソフトおよび各辞書の必用部分は試用期間を設けておき試用期間がすぎたら自動消滅するように設定しておいてもよい。

【0072】また他の方法として前記回線に結合されたサーバー上の音声発生ソフトおよび各辞書にそれぞれ独立に接続のためのゲートを設ける。このゲートを通過するためにはそれぞれ暗証番号を入力することが必用である構造とする。ユーザーは必用部分のみの代金を払い、供給者はこの代金の入金を確認したらユーザーに必用部分の暗証番号を発行する。ユーザーは前記回線を通じてサーバーに接続し音声発生ソフトおよび各辞書の必用部分に前記暗証番号を使い接続する。そして自分のシステムに必用部分をインストールすればよい。

【0073】次にNC(ネットワークコンピューター)的使用の場合について説明する。この場合音声発生ソフトおよび各辞書は公衆電話回線に結合されたサーバー上にのみ存在しているわけであるから、音声を発生させるためにはユーザーの装置から常に前記回線に結合されたサーバーに接続する必用がある。この場合にはサーバーに接続するためのゲートを設けておき、供給者はユーザーの入金を確認したらそのユーザーが前記サーバーに接続可能な状態にするために前記ゲートを通過できるようにすればよい。また前述のように暗証番号を発行しその暗証番号によりゲートを通過できる構造でもよい。

【0074】このゲートを通過可能な状態はユーザーの入金状態により通過不能としてもよい。例えばこのゲートの通過可能状態の期限を設定し、その期限が過ぎた場合通過不能とし、新たに入金があれば再びゲートを通過可能とすればよい。

【0075】他の方法でもよく供給者はユーザーの使用量に応じて請求してもよい。例えば各辞書、各情報テーブルへのアクセスタイムを記録しておき、このアクセスタイムに応じてユーザーの銀行口座から必用金額を引き落とせばよい。

【0076】前述のシステムで作成した音声を録音媒体に記録し、例えば路線バスの車内の案内放送などに使用してもよい。この場合任意の音声を選択することが可能となるのでバスの運行場所、日時などに応じた音声の選択も可能となる。

【0077】本音声発声システムをパソコン通信のチャットのユーザーが使用すれば今まで無味乾燥であったチャットが自分の任意の声を選べるため非常に臨場感のあるものとなる。例えばチャットに参加する人の声を基に音声辞書を作成しておく。そして自分のシステムに前記チャットに参加する人の音声辞書をインストールしておき、その人の発言の時に本システムを稼働すればチャットに参加している人の声でその発言を聞くことが可能となる。本発明の趣旨を逸脱しない範囲で自由に組み合わせればよい。

【0078】

【発明の効果】以上説明したように本発明によれば、人間の声を基にした音声辞書を容易にユーザーに供給することが可能となり、ユーザーのシステムにおいては、ユーザーの任意の音声およびバックグラウンド音響、音声に合致した画像、そして任意音楽が任意の歌声で選択できるため多種多様の音声合成を楽しむことが可能となる。

【図面の簡単な説明】

【図1】実施例の全体のシステムの概要

【図2】ユーザーのシステムの概略フローチャート

【図3】音声辞書の構造

【図4】音声素片の作成のフローチャート

【図5】(a) 標準の音声波のアウトライン

(b) 音声提供者Aの音声波のアウトライン

【図6】(a) 標準の音声波の基本周波数の変化

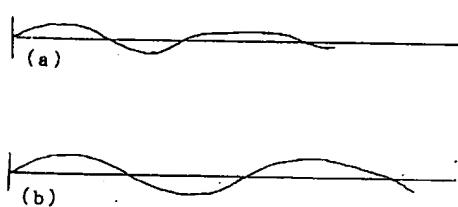
(b) 音声提供者Aの音声波の基本周波数の変化

【図7】環境音響の情報

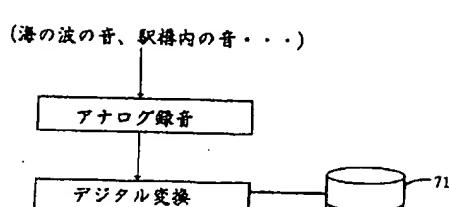
【図8】画像の情報

【図9】方言辞書の構成

【図6】



【図7】



【図10】ユーザーのシステムの詳細

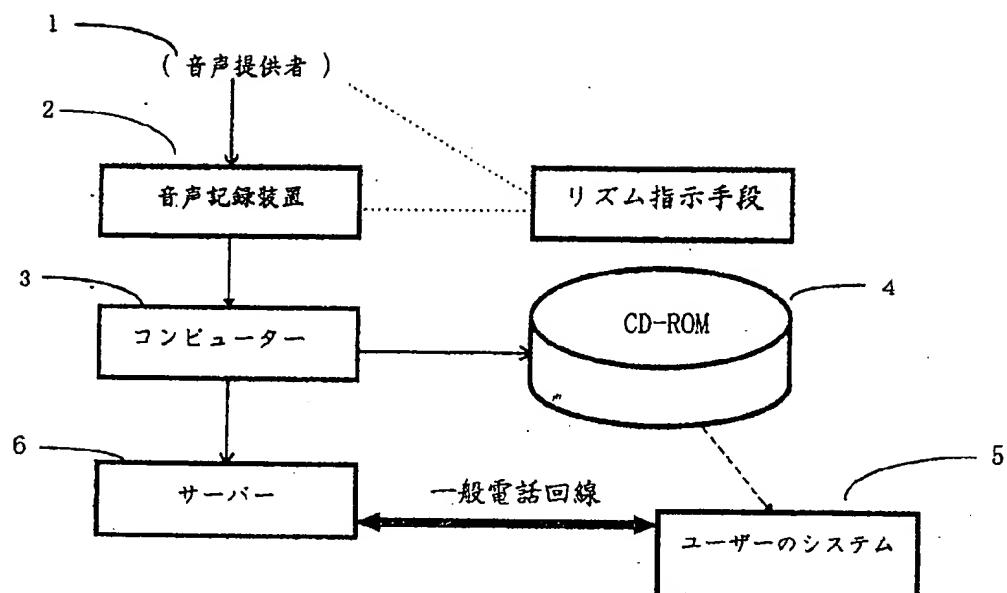
【図11】時間と出力したい音声の種類

【図12】歌声発生のフローチャート

【符号の説明】

1	音声提供者
2	音声記録装置
4	CD-R OM
5	サーバー
21	音声素片辞書
10 22	韻律辞書
23	環境音響情報テーブル
301	音声素片辞書
302	韻律辞書
303	基本周波数部分
304	アクセント部分
305	イントネーション部分
306	時間的特徴
307	音量の特徴
308	方言の特徴
20 309	ポーズの取り方の特徴
310	感情による発声の特徴
408	音素ラベリングテーブル
410	音声素片辞書
71	環境音響情報テーブル
1001	標準辞書
1003 (1)	音声辞書の音声素片部分
1003 (2)	音声辞書の韻律部分
1004	標準化方言辞書
1005	環境音響情報テーブル
30 1006	画像情報テーブル
1010	システム制御部
1011	テキスト解析部
1201	歌情報テーブル
1203	楽器演奏情報テーブル

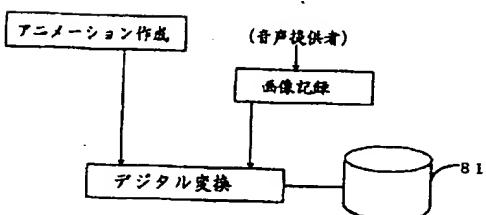
【図1】



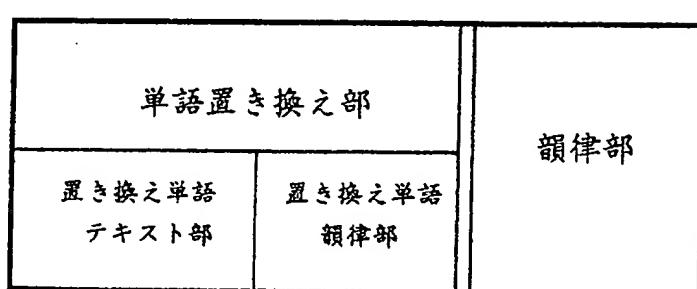
【図3】

音声 素片 辞書	韻律 辞書	基本周波数	102
		アクセント	103
		イントネーション	104
		時間的特徴	105
		音量の特徴	106
		方言的特徴	107
		ボーズの取り方の特徴	108
		感情による発声の特徴	109
			110

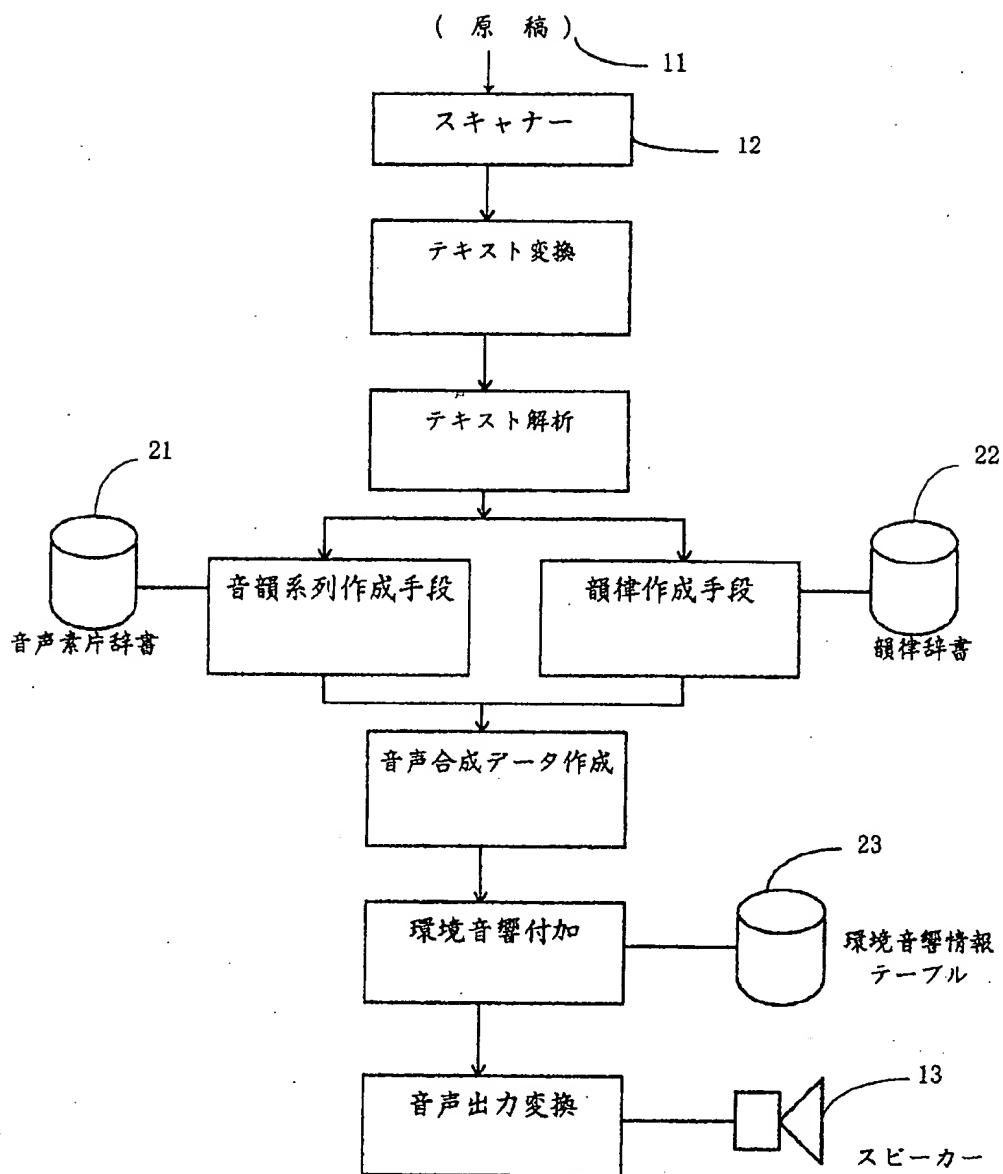
【図8】



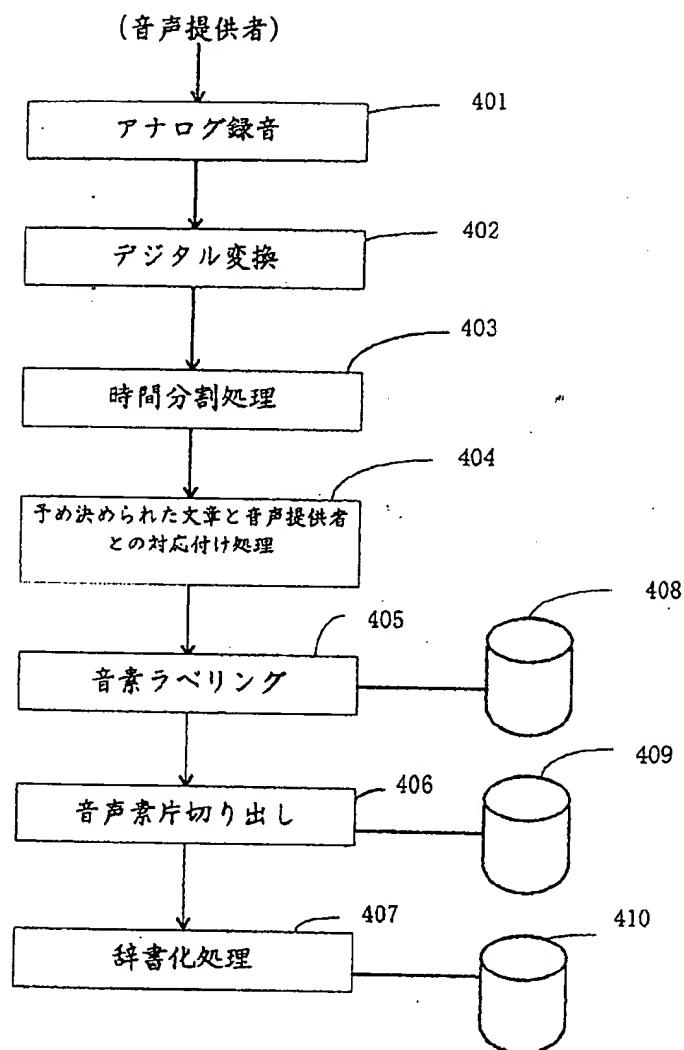
【図9】



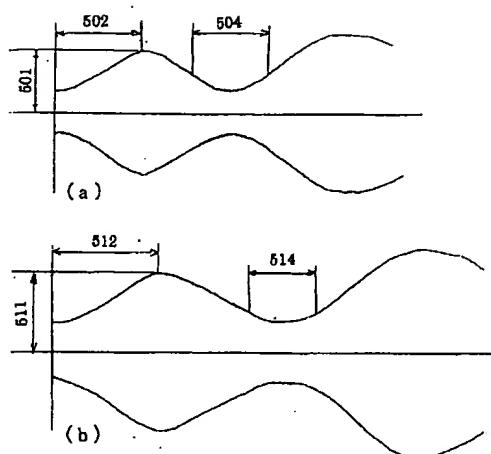
【図2】



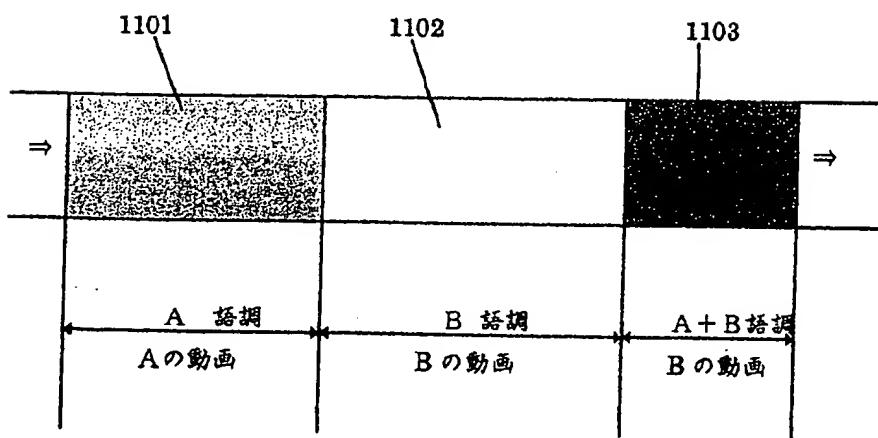
【図4】



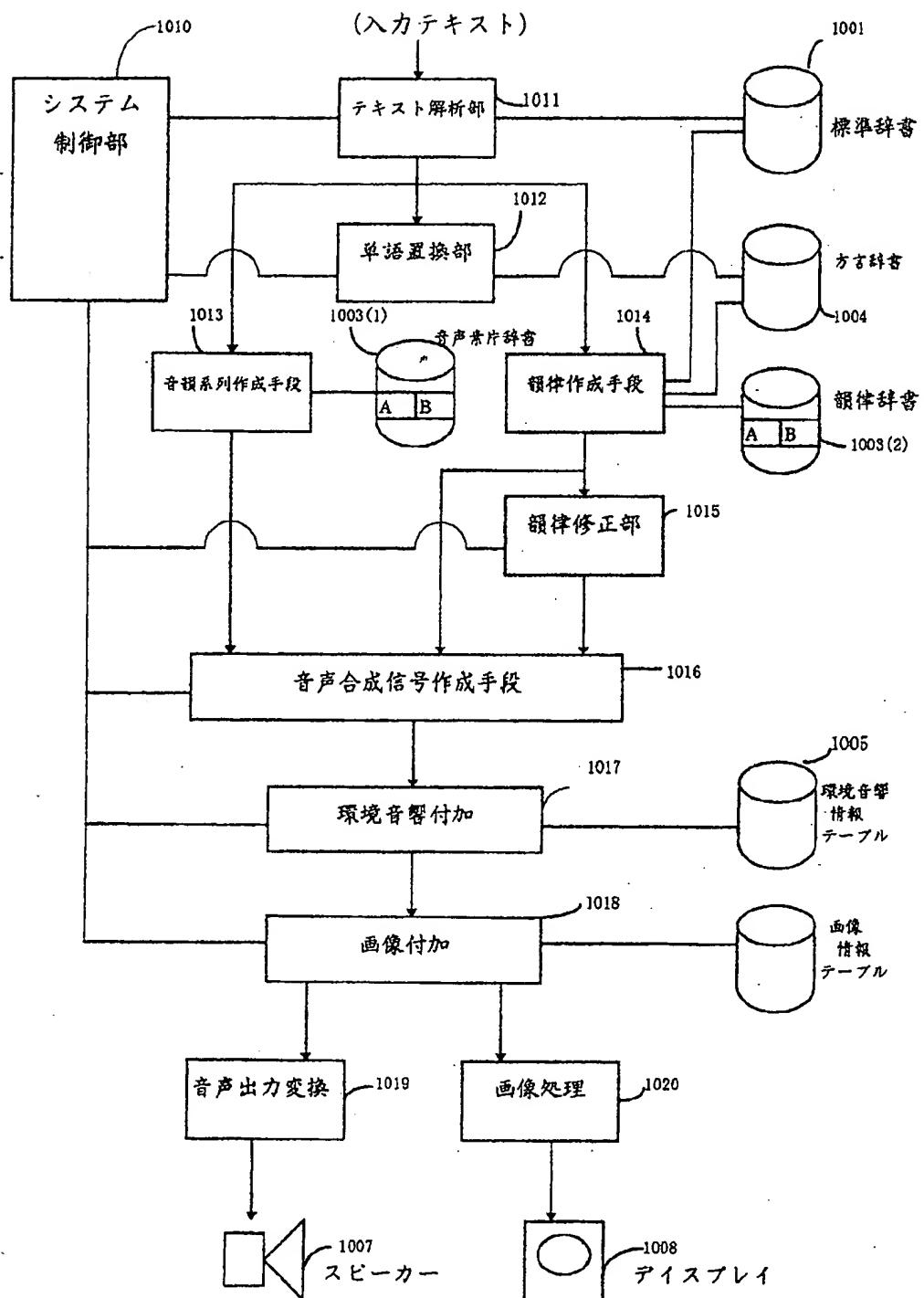
【図5】



【図11】



【図10】



【図12】

